New method for Gamma/Hadron separation in HAWC using neural networks

T. Capistrán[*a], **I. Torres**[a], **L. Altamirano**[a] **and for the HAWC Collaboration**[b]

[a]*Instituto Nacional de Astrofísica, Óptica y Electrónica, Luis Enrique Erro 1, Tonantzintla,*
Puebla 72840, México
[b]*For a complete author list, see www.hawc-observatory.org/collaboration/icrc2015.php.*
Email: tcapistran@inaoep.mx, ibrahim@inaoep.mx, robles@inaoep.mx

The High Altitude Water Cherenkov (HAWC) gamma-ray observatory is located at an altitude of 4100 meters in Sierra Negra, Puebla, Mexico. HAWC is an air shower array of 300 water Cherenkov detectors (WCD's), each with 4 photomultiplier tubes (PMTs). Because the observatory is sensitive to air showers produced by cosmic rays and gamma rays, one of the main tasks in the analysis of gamma-ray sources is gamma/hadron separation for the suppression of the cosmic-ray background. Currently, HAWC uses a method called compactness for the separation. This method divides the data into 10 bins that depend on the number of PMTs in each event, and each bin has its own value cut. In this work we present a new method which depends continuously on the number of PMTs in the event instead of binning, and therefore uses a single cut for gamma/hadron separation. The method uses a Feedforward Multilayer Perceptron net (MLP) fed with five characteristics of the air shower to create a single output value. We used simulated cosmic-ray and gamma-ray events to find the optimal cut and then applied the technique to data from the Crab Nebula. This new method is tuned on MC and predicts better gamma/hadron separation than the existing one. Preliminary tests on the Crab data are consistent with such an improvement, but in future work it needs to be compared with the full implementation of compactness with selection criteria tuned for each of the data bins.

---

[*]Speaker.

arXiv:1508.04370v1 [astro-ph.IM] 18 Aug 2015

1. Introduction

The High Altitude Water Cherenkov (HAWC) gamma-ray observatory is composed of 300 water Cherenkov detector (WCD). On the bottom of each WCD there are 4 photomultiplier tubes (PMTs) that detect the Cherenkov light. This light is produced by secondary particles in air shower generated by the interaction between atmosphere and primary particle (as for example gammas rays, protons, among other particles). The rate of cosmic rays (CR) is bigger than the gamma rays (GR) so it is critical to find a technique to remove the CR without losing the signals of GR.

Currently, HAWC has a method called compactness for distinguishing those primary particles. For doing this, the data is divided into 10 bins (see Table 1) depending on $nHit$, that is the number of PMTs that have a signal in the event. The compactness depends upon the charge distribution deposited by the secondary particles of the shower on PMTs of the array. In this work, a new method is presented, using a Neural Network (NN) for the gamma/hadron separation without dividing the data into bins. Five characteristics are computed for feeding a NN that computes a value (θ_{NN}) to distinguish between CR and GR. Another method in development can be found in [1].

2. Training stage

The NN used in this work is a Feedforward Multilayer Perceptron [2]. For a correct evaluation, the NN must pass two stages, training and testing. In the training stage the aim is to minimize the classification error. First, the values of characteristic input are calculated and a training MC data set is selected. The architecture is defined as 5-5-5-1 (Figure 1a), the first layer has 5 neurons because the NN need 5 characteristics as input[1], one neuron in the last layer because the network needs to recognize only two types of particle. Different architectures of NN were tested but the learning curves were similar. In the use of NN the recommended number of total layers should be $N-1$ where N is the number of input variables [3], in our case $N = 5$ so the simple structure (5-5-5-1) was chosen to save computing time. The learning method used was stochastic minimization and took 500 epochs for a asymptotic behavior in the error of the output.

In Figure 1b is shown the histogram of the output for the NN. The majority of the events produced by GR are close to value 1 and CR to 0. Finding the optimal cut in this variable will allow us to separate between different types of primary particles. This threshold value is defined as θ_{NN}.

The Q factor is defined as $\varepsilon_{gamma} \, / \, \sqrt{\varepsilon_{hadron}}$ where ε_{gamma} is the fraction of gamma events that are classified correctly, also called gamma efficiency, and ε_{hadron} is the hadron events that are classified as gamma events, also called hadron efficiency. The Q value estimates the factor by which the significance will be increased by the classification. Figure 2a shows the Q factor and the θ_{NN} value, where it can be seen that the highest value of Q corresponds to a value around $\theta_{NN} = 0.98$. The receiver operating characteristic (ROC) curve is useful for comparing classifiers and visualizing their performance [4]. From the ROC curve we can see that by using $\theta_{NN} = 0.96$ we increase the gamma efficiency, even if we have a bit lower Q Factor with this cut (see Table 2).

---

[1]Each input is normalized with respect to maximum value of each feature.

bin	nHit min	nHit max	θ_c
-1	30	54	-
0	55	87	4.6
1	88	138	6.3
2	139	216	9.8
3	217	323	12.7
4	324	457	17.6
5	458	606	19.5
6	607	754	18.5
7	755	889	17.1
8	890	1000	15.0
9	1001	1200	12.4

Table 1: nHit range and gamma/hadron cut in each bin for HAWC-300, θ_c is the compactness cut value.
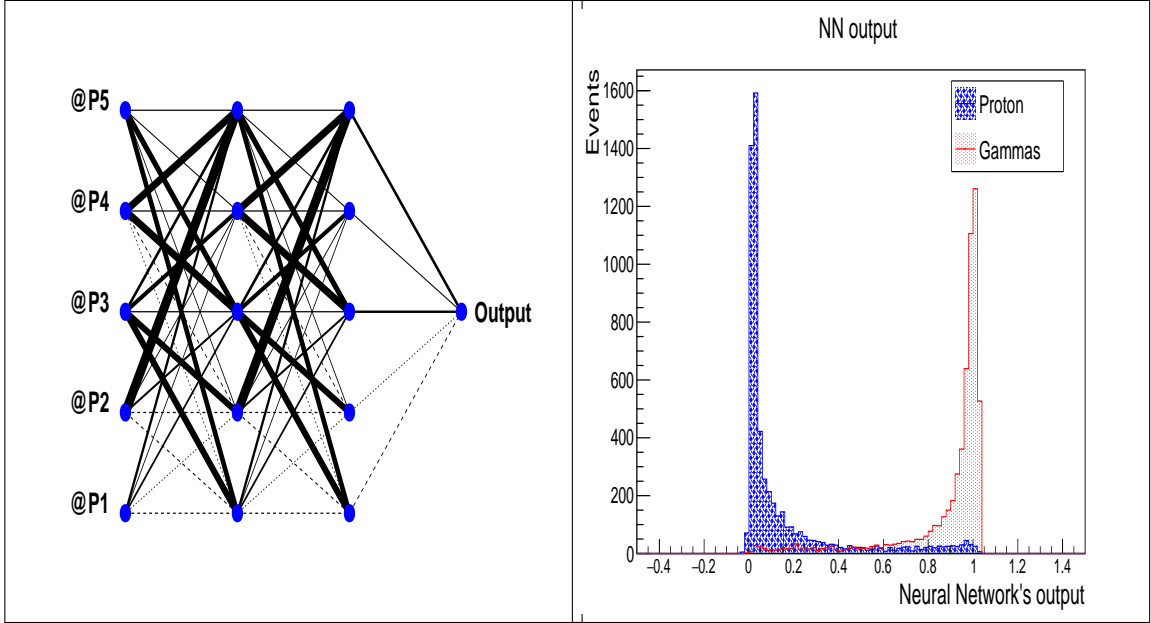

(a) Architecture of NN.　　　　　　　　　　　　　(b) Outputs of NN.

Figure 1: In (a) is shown the architecture of NN with 5 neurons as inputs, two hidden layers with 5 neurons and one neuron as output. The width of each connection line between neurons is proportional to the weight of the NN. In (b) is shown the outputs of the NN for gammas and hadrons in the learning stage. The majority of gamma events have an output close to one, and protons are close to 0.

2.1 Choice of characteristic inputs

The main idea is to use the morphological differences of the charge distribution in the PMTs for the two type of primary particles. In event produced by gammas the PMTs close to the core of the shower have the biggest signals and the charge distribution is characterized by a compact and
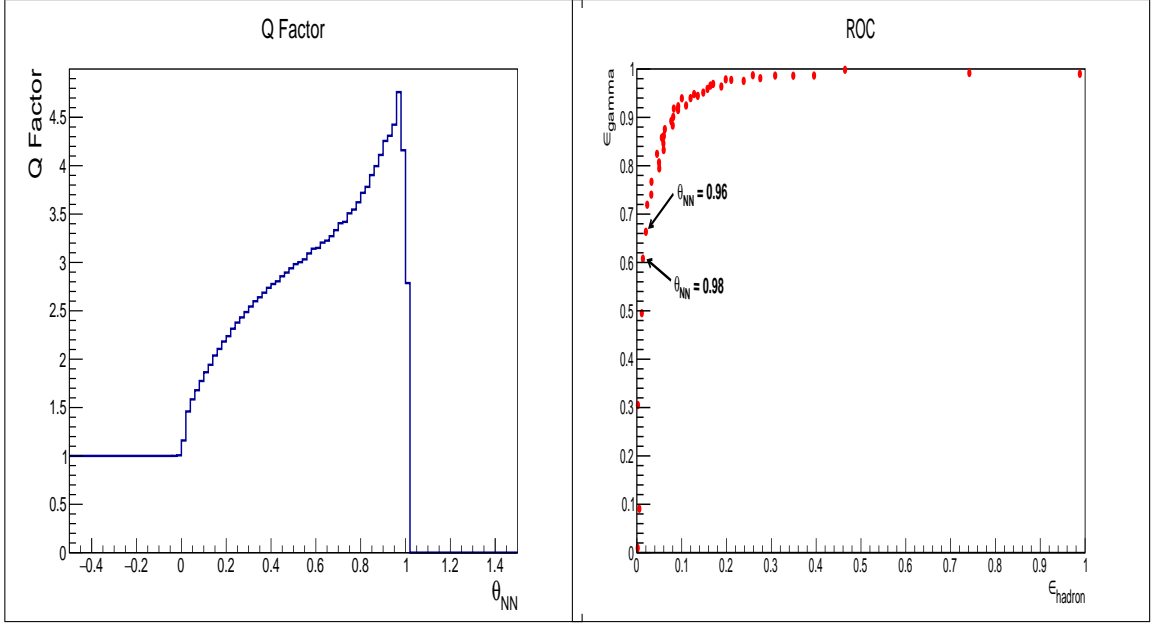
(a) Q Factor.   (b) ROC curves.

Figure 2: In (a) is shown the Q Factor of NN's outputs. The largest Q factor is at 4.76 when the output threshold is around 0.98. In (b) are shown the ROC for the NN. The θ_{NN} corresponding to the ε_{gamma} between 0.6 and 0.7 could been used, at a loss of some Q value.

ε_{NN}	ε_{gamma}	ε_{hadron}	Q Factor
0.94	0.713	0.028	4.309
0.96	0.666	0.024	4.424
0.98	0.604	0.019	4.761
1.00	0.495	0.011	4.160
1.02	0.306	0.005	2.787

Table 2: Values for gamma and hadron efficiency close to the maximum value of Q factor. Here, for completeness, we include the bin -1 from Table 1, even thought the bin is not used in the compactness analysis.

smooth profile. But in the case of hadrons, PMTs with high charge can be far away from the core and the charge distribution is not compact.

- The first feature we include is the number of PMTs with at least one photoelectron (PE) because it is directly related to the energy of the primary. We need our NN to distinguish independently of the energy of the CR or GR. This replaces the nHit binning used with the compactness cut (P1=nHit).

- *DisMax* (P2) that is the largest distance between any of the pair of tubes passing the next selection: first all the PMTs in the event are sorted by their PEs detected and we summed this value for each PMT from higher to lower until the sum is less that $(SumPE - MaxPE)k(nHit)$,

where *MaxPE* is the number of PEs in any PMT in the event, and "k" is a factor that depends linearly of nHit, the PMTs involved in that sum are the selected ones. This input involves the distance of the PMTs with biggest charge detected and its distance because we suppose that for gammas all the PMTs with high PE are neighbors and this *DisMax* should be small.

- *P*3 this feature is associated with the integral of the radial density where the hadron shower dominates gamma shower [5] defined as:

$$Log_{10}\left(\frac{nHit}{\sum_n PE_i * R_{PE_i}}\right) \qquad where\ R_{PE_i}\ >\ 30\ m$$

Here PE_i is the charge in the PMT_i, R_{PE_i} is the distance in meters between the PMT_i and position of the reconstructed shower center (core).

- *P*4 is defined as $CxPE_{30}/MaxPE$, where $CxPE_{30}$ is the maximum charge outside a exclusion radius of 30 m in the event. For protons one expects to often see charge localized high charge deposition far from the core, so P4 can approach 1 for protons. On the other hand, gammas usually have a value near 0 because most of their charge is deposited near the core.

- *P*5 is related to the difference between the maximum charge outside and inside the exclusion radius weighted with the distance to the core.

$$P5 = Log_{10}(|CxPE_{30} * R_{CxPE_{30}} - PE_{maxint} * R_{PE_{maxint}}|)$$
$$where\ R_{CxPE_{30}}\ >\ 30\ m\ and\ R_{PE_{maxint}}\ <\ 30\ m$$

2.2 Training data set

The simulated events were generated by using CORSIKA program in the energy range [0.005,100] TeV with a flat spectrum and zenith angle $[0,75]^o$. The performance and response of the array were computed using the HAWC official software.

For the training stage, the network need two data sets, one for gamma and other for hadron. We defined a target value of 1 for gamma ray event and 0 for hadron event. In this work we only use protons as hadrons because protons constitute nearly 99% of the CR. The conditions for selecting training the events for each set are:

- The event is well reconstructed.

- The difference between the core reconstruction and simulation does exceed 5 m.

- The core falls inside the HAWC array
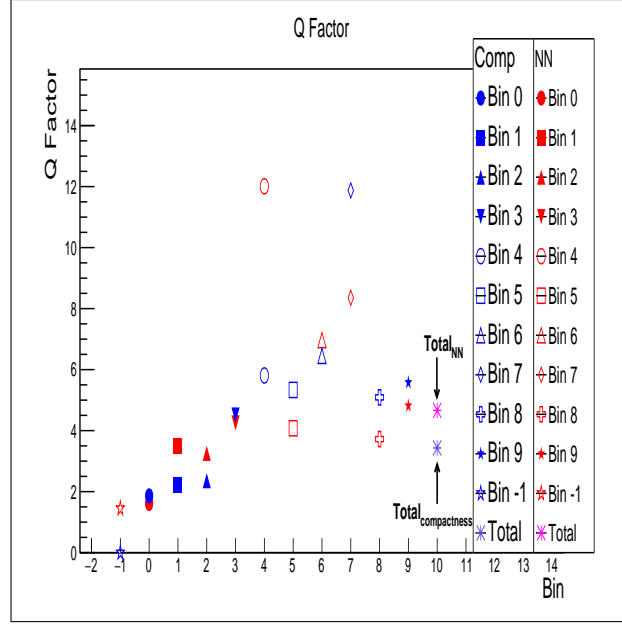
- Event with nHit between 30 and 1200.

Figure 3: The Q factor is calculated for each bin and the total (bin 0 to 9) with $\theta_{NN} = 0.96$. This shows that for the Q factor in some bins, the NN is better than compactness but for others does not. Using the total bin we got 60% and 53% in gamma efficiency for NN and compactness respectively.

3. Testing stage

3.1 Simulation

In this stage we use the same criteria described above for selecting the events for the training data set which consists of new simulated events independent of the training set. For comparing the two methods we use events with $55 < nHit < 1200$, that correspond to bin 0 up to bin 9, i.e. we are not using the bin -1. In this comparison we will simply weight all events equally, without the optimal weighting for events in each bin used in [6] the Crab analysis. However, we do apply the compactness cuts of Table 1 for each nHit bin to compare performance of the NN and compactness. The bin called "total" is computed using all events from bin 0 to bin 9. The results are shown in Figure 3 where we can see that for the Q Factor the NN has a better result than using the compactness method.

The total value of Q Factor , gamma efficiency and hadron efficiency of each separation methods (compactness and NN) is shown in the Table 3. The NN improves on the compactness method. The gamma efficiency increased by 13% and the hadron efficiency decreased 30%, so the Q factor increased by 35%.

3.2 Data

Another way to compare the different performance of the compactness and the NN is using HAWC data. We chose a set of well reconstructed events within $\pm 6°$ of the Crab Nebula. We have two methods (NKG and Gauss) for reconstructing the core position, but only Gauss was used in

Parameter	NN	compactness	Increase (%)
Q Factor	4.663	3.432	35.889
gamma efficiency	0.606	0.536	13.129
hadron efficiency	0.017	0.024	-30.693

Table 3: Difference between methods with simulation.

θ_c	NKG	Gauss
10.0	3.4706	4.4649
12.0	4.3142	4.4703
14.0	5.2777	4.6895
16.0	3.9327	4.3406
18.0	4.3170	4.3613

Table 4: Significance using the compactness variable with a single cut value for all bins.

θ_{NN}	NKG	Gauss
0.92	5.8842	4.9889
0.94	5.7042	5.4144
0.96	5.9217	5.5096
0.98	3.7534	4.6703
1.00	4.0977	3.1792

Table 5: Significance using NN Vs NN threshold.

training the NN. A well-behaved event should have a similar core position for either method. In the case of using compactness we use a very simple analysis [7] and apply a cut of θ_c that varies from 10 to 18 but is applied to all nHit bins. For technical reasons we were not able to apply the bin-dependent cuts of Table 1 to the Crab data, so this constitutes a preliminary comparison of NN and compactness on the Crab data. The results are shown in Table 4. In the case of NN method, the maps are obtained by varying θ_{NN} from 0.92 to 1.0 (see Table 5).

The highest values of significance from Tables 4 and 5 are placed in Table 6 and the increase with respect to the compactness method is computed. The results show that the NN is better than compactness in this preliminary comparison, consistent with expectations from MC. With the NKG method, the increase is 12 %, and with Gauss method is 17 %, not surprising since the NN learnt with events whose core reconstruction was done with Gauss method.

4. Conclusion

In this work, we propose a new method for gamma/hadron separation that used a Multilayer Perceptron fed with 5 characteristics. The NN's output is continuous and has a value targeting 1

Method	NKG	Gauss
compactness	5.2777	4.6895
NN	5.9217	5.5096
Increase (%)	12.202	17.488

Table 6: Difference between methods with data.

for gamma events and 0 for hadron events. In the analysis, we found an optimal cut value for the NN output $\theta_{NN} = 0.96$. With this value the NN has better performance than compactness. The Q Factor increases approximately 36%, because the gamma efficiency increased about 13% and a decrease of 30% in hadron efficiency.

In the case of Crab data we also obtained a better significance using NN instead of a simplified version of compactness where the compactness cut was constrained to be the same for all nHit bins. In future work we will compare with the full compactness implementation.

Acknowledgments

References

[1] **HAWC** Collaboration, Z. Hampel, *Gamma-Hadron Separation using Pairwise Compactness Method with HAWC*, in *Proc. 34th ICRC*, (The Hague, The Netherlands), August, 2015.

[2] Christophe Delaere, *class TMultiLayerPerceptron*, 2013 (accessed January, 2014).
http://root.cern.ch/root/html/TMultiLayerPerceptron.html.

[3] C. M. BISHOP, *Neural Networks for Pattern Recognition*. Oxford University Press, Inc. New York, NY, USA, 1995.

[4] T. Fawcett, *An introduction to ROC analysis*, *Pattern Recognition Letters* **27** (2006) 861-874.

[5] V. Grabski, L. Nellen, and A. Chilingarian, *Gamma/hadron separation study for the HAWC detector on the basis of the multidimensional feature space using non parametric approach*, in *Proc. 32th ICRC*, vol. 9, (Beijing, China), p. 207, August, 2011.

[6] **HAWC** Collaboration, F. S. Greus, *Observations of the Crab Nebula with Early HAWC Data*, in *Proc. 34th ICRC*, (The Hague, The Netherlands), August, 2015.

[7] **HAWC** Collaboration, J. Pretz, *Enabling Fast Efficient Crab Verification*, tech. rep., internal Document, 2014.